puter and information ethics; *Severson*: Principles of information ethics; *Smith*: Information ethics.).

Sokasodnak az információs etikával foglalkozó konferenciák is (pl. az UNESCO rendezésében az Infoethics 1997-böl, 1998-ból és 2000-ből). Az IFLA ugyancsak fokozza a téma iránti érdeklödését, aminek külön bizottság felállítása lett a következménye.

Nagy a divatja az általában könyvtáros etikai kódexeknek (pl. Hollandia, 2002; Litvánia, 1999; Oroszország, 2002). Az International Federation of Information Processing információs etikai kódexet készít elő. A nagy-britanniai CILIP az LA régebbi etikai kódexe helyett újat akar elfogadni.

Az etikai kódexek kívánatos tartalmát illetően több elképzelés van forgalomban. Így az International Center for Information Ethics szerint az információs etika keretébe beletartozik a mediális, a számítógépi, a biológiai, a könyvtári, az üzleti információs és az internetes etika.

Az információval hivatásszerüen foglalkozó szakemberek számára föként a következö területeken adódnak etikai teendök (vö. *Thomas Froelich* jelentésével az UNESCO General Information Programme c. dokumentumához 1997-böl):

- a hozzáférés biztosítása mindenkinek,
- az információ-elöállítás etikai szempontjainak betartása,
- az információgyüjtés és -feldolgozás etikai vonzataihoz való igazodás,
- az információhasználat közben érvényesítendö etikai normák.

Az emberiség jövője nem kis mértékben múlik az információval foglalkozó szakemberek fentieknek megfelelő etikus magatartásán.

/SOSINSKA-KALATA, Barbara: Etyka w nauce o informacji. = Bibliotekarz, 9. sz. 2003. p. 3–10./

(Futala Tibor)

Ugorj magasabbra: webhelyek rangsora a Google-nál

Mitől lesz egy weblap a keresőmotorok sztárja? *Lisa Zhao*, az Illinoisi Egyetem katalogizálási osztályának munkatársa azt vizsgálta, hogy melyek azok a jellemzők, amelyek egy webcímet a visszakeresés élére katapultálnak.

Az OCLC (Online Computer Library Center) adatai beszédesen mutatják a hálózati helyek ugrásszerű növekedését. A nyilvános (döntően szabad hozzáférésű) weblapok száma 1998–2002 között 1 457 000-ről 3 080 000-re emelkedett. A magánjellegű (közönség számára korlátozott hozzáférésű) webhelyek száma 315 000-ről 2 489 000 lett. Az ideiglenes (triviális, nem közérdeklődésre számot tartó) URL-ek száma pedig 864 000-ről 3 143 000-re nőtt a jelzett négyéves időszakban. Az információk sebesen buzogó áradata, és a használók gyorsan lankadó figyelme miatt a tartalomszolgáltatók számára húsbavágó kérdés lett, hogy honlapjuk a keresőmotorok találati halmazának élbolyába kerüljön. A világhálón szörfölők keresési szokásairól végzett kutatások egyre lehangolóbb eredményeket mutatnak. A használók fele csupán egyetlen szót ír be a keresőcsíkba. Döntő többségük a találati halmaz töredékét sem nézi meg – lustaságból vagy türelmetlenségből. *Bernardo A. Huberman* szerint egy adott hálócímen az átfutott lapok száma három, a keresők zöme csupán egyetlen oldalt hív elő, és az első húsz találatnál tovább nem tekint. Megbízható becslések szerint az átlagos használó türelme négy és tíz másodperc között van.

Ma a legismertebb és leggyakrabban használt keresőgép a *Google,* amely egy 2003. szeptember 3-i adat alapján 3,3 milliárd weblapot indexelt, s naponta 70 millió keresést fogadott. (2005. január 1-jei adat: a Google 8 milliárd lapot gyüjtött be, és egy nap alatt 200 millió keresést kezelt. – A ref.) A weblapok rangsorolására szolgáló eljárásrendszer minden keresőmotor féltett titka. Bonyolult algoritmusok sora, számos tényező kombinálása eredményezi egy adott hálóhely súlyozását. Egyetlen nyilvánosságra hozott oldalpozicionáló algoritmust ismerünk: a Google *PageRank (PR)* elnevezésű alkalmazását. (Magyarul talán a PR lehetne *P*ozíció a *R*angsorban, így a PR rövidítést meg tudjuk tartani. – A ref.) A PR elsődleges fajsúlyos tényezője a link, a kapcsolat: az, hogy hová kötünk be linket, egy szavazatot jelent az adott weblapnak. A PR-ről első ránézésre annyi kiderül, hogy nemcsak az számít, hogy egy bekapcsolt link saját PR-je

milyen, hanem az is, hogy az adott másik honlaphoz milyen PR-rel rendelkező linkek mutatnak. Mennél magasabb a hozzánk bekötő weblapok PR-je, annál magasabb a saját weblapunk PR-je is. Hangsúlyozni kell azonban, hogy a PR csak egyike az alkalmazott algoritmusok sorának, és a weblapok tényleges pozíciója a találati halmazban még számos egyéb – száznál is több – tényezőtől függ. Íme a PR-t kifejező matematikai képlet:

PR (A) = (1-d) + d(PR(t1)/C(t1)+...+PR(tn)/C(tn)),
ahol:
A – saját webhelyünk;
PR(A) – Pozíció a Rangsorban *(PR)*, amelyet webhelyünk *(A)* kap, amikor egy másik webhely *(t1)* hozzánk linkel;
d – súlyozási faktor (damping factor), amelynek állandó értéke 0,85;
t1 – másik, hozzánk linket létesítő webhely;
tn – hozzánk linket létesítő webhelyek összessége,
C(t1) – hozzánk linket létesítő webhely (t1) saját összes linkjének száma;
PR(t1) – hozzánk linket létesítő webhely (t1) rangsori pozíciója, *PR*-je.

A *(tn)* faktor tehát azon weblapok összessége, amelyek linket létesítenek hozzánk. *Phil Bradley* egyes keresőmotoroknál azt az elvet véli felfedezni, hogy ha sok ember kapcsolódik egy helyhez, akkor az „jobb" hely, mint egy olyan, amelyhez kevés ember létesít linket. Itt persze az a gyenge pont – mutat rá Bradley –, hogy nyilván az új lapokhoz kevesebb link kapcsolódik, mint a régebbiekhez.

A PR mellett, ahogy már szó volt róla, a Google más jellemzőket is figyelembe vesz. A keresőkifejezések, kulcsszavak elhelyezkedése a weblapon és azok gyakorisága két további tényező. Fontosságban előbbre áll, ha a lap tetejénél van egy kulcsszó, pl. szalagcímben, a headerben vagy az első bekezdésekben. Nem elhanyagolható sorrendi elem az sem, ha a keresőkérdés a HTML címjelölőjében, a TITLE-tagben szerepel. A kulcsszavak gyakorisága a sűrűségi tényező révén szól bele a rangsorba – a sűrűség a keresőszavak gyakorisági aránya az adott weboldalon lévő szavak összességéhez képest. A szavak túl gyakori szereplése, túlzott dominanciája azonban visszájára fordul: az ideális sűrűség mindössze 6–10%. A korábban visszaélésekre alkalmat adó jelölőket, a *meta description*t és a *meta keywords*t ma már a robotok nem veszik tekintetbe, és a rangsorolásnál ezek semmit sem számítanak.

A weblap-rangsorolási tesztet a szerző 2002. november 13. és 2003. január 15. között tíz héten át heti egy alkalommal, ugyanazzal a keresőkérdéssel végezte a Google-ban. Empirikus kutatásában arra volt kíváncsi, hogy mi az első húsz találat, s a halmaz webhelyeinek jellemzői milyen módon állnak összefüggésben rangsorbéli helyezésükkel. A következő tényezőket vizsgálta: (1) a PR súlya; (2) a webhely népszerűsége; (3) a keresőszavak (kulcsszavak) száma és sűrűsége; (4) a keresőszavak elhelyezkedése a honlaptartalom címében, a HTML-címjelölőben és az URL-ben; (5) a domén szerkezete, hierarchiája. A teszt a Google egyszerű keresőlapjával futott, a keresés kulcsszavai pedig: *cataloging department* (katalogizáló osztály) – de nem kifejezésként idézőjelben, hanem szavankénti keresésként.

Eredmények és elemzés

A tízhetes teszt alatt 24 különböző webhely került a húszas listába. Csak három pozíciót, az elsőt, a negyediket és a hetediket foglalta el ugyanaz a weblap végig a teszt alatt. A többi pozíció birtokosa rendre változott. Az első 13 intézmény a teszt tíz hete alatt végig megjelent a top twentyben. Ezek nem változtatták a helyezésüket ± három hellyel többet hétről hétre, s egyik sem került a 14. helynél hátrább. Ellenben a maradék tíz webhely változtatta a helyét a 14–24. pozíciókon. A szerző szerint az, hogy csupán 24 weblap jelent meg a húszas listán, azt mutatja, hogy a keresőmotor rangsorolása stabil alapokon nyugszik.

A Google rangsorolási algoritmusa, a PR

A fenti intézmények PR-jét a Google tool bar egyik letöltött függvényével mérték. A legtöbb weblap PR mutatója 4 és 7 közé esett. A húszas lista egyes tagjai azonos PR-értéket mutattak. Az első tíznél a PR 5 vagy 6, míg a másik tíznél 2 és 7 között állt. Általában megállapítható, hogy a tízes listától lefelé a PR a helyezési sorrendtől függően esni kezd annak ellenére, hogy a sereghajtók közül néhány weblap magasabb PR-t mutatott.

A weblap népszerűsége

Egy weblap népszerűségének fokmérője, hogy hány weblap köt linket hozzá. A Google PR nem tükrözi az adott webhely népszerűségét, mivel más tényezőket is figyelembe vesz. Tehát amikor sok linkkapcsolat mutat egy webhelyre, annak lehet magas PR-je, de lehet alacsony is. Alacsony PR

nem feltétlenül a kisebb népszerűség jele. (A weblap népszerűségi felmérése hasonló a nyomtatott folyóiratok impakt faktor rangsorához. A folyóirat impakt faktora annak a gyakoriságnak a mértéke, amellyel az adott folyóirat átlagos cikkét idézték egy adott évben. Az impakt faktor segít a folyóirat fontosságának kiértékelésében – különösen az adott diszciplínán belül más folyóiratokkal való összevetésben.) A jelen vizsgálatban úgy állapították meg az adott webhelyhez egyénileg linkkel kötődő weblapok számát, hogy a Google-keresésben a site-ok URL-jére kerestek (ennek szintaxisa: link:<url>). Az egyes számú hely a népszerűségi listán magasan kiemelkedik a mezőnyből. Az 1. helyet végig ugyanaz az intézmény *(University of Virginia Library, Cataloging Services)* foglalta el a teszt folyamán. A népszerűség és a rangsor összefüggéséről annyi megállapítható, hogy a Google bizonyos mértékig számításba veszi a népszerűséget, ám – helyesen – azt nem értékeli túl, tudván tudva, hogy az újabb webhelyek eleve kevésbé népszerűek, hiszen kevesebb linkjük van, mint a régieknek.

Kulcsszavak száma és sűrűsége

Mivel a *cataloging department* keresőkérdést idézőjelek nélkül – tehát nem kifejezésként – írták be a Google-keresőbe, a tesztkeresés külön és együttes találatokat is visszaadott. Megszámolták, hogy a *cataloging* és a *department* kulcsszók hányszor szerepelnek mindegyik webhely nyitólapján, kiszámolták a kulcsszósűrűségi indikátort, és összevetették ezeket a számokat a rangsorral. Átlagban a honlapokon a legkisebb kulcsszószám 2, a legnagyobb 30. A kulcsszósűrűséget úgy számítják ki, hogy összevetik a kulcsszavak számát a honlap összes szószámával. A kulcsszósűrűség a legtöbb honlapnál a 4–13% közötti tartományba esett.

A kulcsszavak nagy száma nem mindig jelez magas kulcsszósűrűséget egyúttal. Pl. az első héten a 18. helyen lévő intézmény honlapján a legnagyobb volt a kulcsszószám, ám a kulcsszósűrűségben hátrább állt. Nem az első helyen álló webhelynek volt a legnagyobb kulcsszószáma, sem kulcsszósűrűsége. Ugyanakkor az utolsó helyen lévő webhelynek sem volt a legkisebb kulcsszószáma vagy kulcsszósűrűsége. Összefoglalva: sem a keresőszavak (kulcsszók) száma, sem azok sűrűsége nem jelez szignifikáns összefüggést a halmazbéli rangsorral. Ám a fordítottja igaz: ha a webhely rangsorban visszaesik, akkor a megfelelő keresőszavak sűrűsége is csökken.

Kulcsszópozíciók

A teszt annak vizsgálatára is kitért, hogy a *cataloging* és *department* keresőszavak, kulcsszavak megjelentek-e három helyen: a honlap tartalmi címében (title), a HTML cím-tagben és az URL-ben, és ezek rangsorbeli helyezésre gyakorolt hatását elemezte.

Kulcsszavak a tartalom címében

A tartalmi cím a szövegtestben a honlap tetején elhelyezkedő cím, a HTML heading-tagek <H1>...</H1> közötti szöveg. Mind a huszonnégy intézmény vagy a *cataloging department*et vagy a *cataloging* szót tartalmi címében megjelöli. 18 webhely tartalmazta a *cataloging department* szavakat a tartalmi címben, hat intézménynél pedig a *cataloging* szerepel.

Kulcsszó a HTML title-tagben

A HTML egyik kötelező jelölőcímkéje, tagje a <TITLE>...</TITLE>. Ez tartalmazza a dokumentum címét. Az itt szereplő cím nem jelenik meg a böngészőablakban, csupán annak címsorában. *Danny Sullivan* szerint azok a helyek, ahol a keresőkifejezések szerepelnek a TITLE-tagben, nagyobb relevanciát mutatnak azoknál, ahol nincsenek a TITLE-tagben. Ennek ellentmond, hogy a teszt során az első tíz intézménynél nem szerepel mindkét keresőszó a TITLE-tagben, míg a 11–20-as mezőnyben mindkét kulcsszó ott van. Röviden, a kulcsszavak jelenléte a HTML cím-tagben nem igazán van hatással a Google rangsorolásában.

Kulcsszó az URL-ben

A keresési eredményt feltehetőleg befolyásoló másik kulcsszó-elhelyezkedés az URL. Lisa Zhao a *cataloging, catalog, cat, dept* szavak használatát kutatta. Az első tizennégy intézmény valamilyen formában jelzi a kulcsszót a hálózati címében, s az megállapítható, hogy az URL-ben szereplő kulcsszavak hatással vannak a sorrendre.

A webhelyek doménszerkezete

Az utolsó vizsgált szempont azt volt hivatott kimutatni, hogy a webhely elhelyezkedése a saját doménhierarchián belül mennyire befolyásolta a Google-rangsorán belüli pozíciót. A huszonnégy intézmény, amely a tízhetes teszt alatt az első húsz helyet elfoglalta, kilenc típust képvisel URL-szerkezet szempontjából:

1. könyvtári domén/katalogizáló osztály (7 ilyen intézmény van – az 1. szint, a katalogizáló osztály weblapja a domén alatt egy szinttel van)

2. könyvtári domén/osztályok/katalogizálás (6 ilyen intézmény van – ugyancsak ide tartozik a könyvtári domén/technikai szolgálatok/katalogizálási osztály – 2. szint)
3. könyvtári domén/könyvtárak/katalogizálás (1 ilyen intézmény – 2. szint)
4. könyvtári domén/osztályok/részlegek/katalogizálás (1 intézmény – 3. szint)
5. könyvtári domén/könyvtárak/egységek/osztályok/katalogizálás (1 ilyen intézmény – 4. szint)
6. intézményi domén/~katalogizálási osztály (1 ilyen intézmény – 1. szint)
7. intézményi domén/~technikai szolgálatok/katalogizálás (1 ilyen intézmény – 2. szint)
8. intézményi domén/könyvtár/katalogizálási osztály (2 ilyen intézmény – 2. szint)
9. intézményi domén/könyvtár/osztályok/katalogizálási osztály (4 ilyen intézmény – 3. szint)

Egyik osztály sem birtokolt saját domént. A teszt során végig az első, a negyedik és a hetedik helyen lévő intézmények a könyvtári domén/katalogizálási osztály típusú doménszerkezetbe tartoztak. Az első hét webhely könyvtári domén alatt helyezkedik el. Az első tizenegy 2. szintnél nem áll lejjebb a doménhierarchiában. Megállapítható, hogy mennél mélyebben állt az intézményi doménhierarchiában a webhely, annál hátrább állt a találati halmaz rangsorában.

* * *

Összegzésül: a szerző szerint a Google-találati halmazban előkelőbb helyezéshez, magasabb pozíció eléréséhez a tényezők bonyolult együtthatása játszik közre. A nyolc vizsgált tényező közül öt – a Google PR, a webhely népszerűsége, a kulcsszavak sűrűsége a honlapon, a kulcsszavak az URL-ben, a doménszint – fontosabbnak bizonyult, mint a többi. Ám a titkos rangsoroló algoritmusok és a keresőmotorok szakadatlan fejlesztése miatt még sok empirikus-tudományos kutatást kell végezni ahhoz, hogy pontosan megállapíthassuk, mitől kerül egy webhely a találati halmaz csúcsára.

/ZHAO, Lisa: Jump higher: Analyzing web-site rank in Google. = Information Technology and Libraries, 23 köt. 3. sz. 2004. p. 108–118./

(Bánhegyi Zsolt)

Az Ingenta megújult könyvtári szolgáltatásai

A múlt

Az *Ingenta* a BIDS (Bath Information and Data Services) 1991-ben indított, a JISC által finanszírozott szolgáltatásához kapcsolódóan született. A BIDS-ről azt állítják, hogy a világon az első nemzeti szintű szolgáltatás volt, és kereskedelmi adatbázisok hálózati elérését biztosította az Egyesült Királyság oktatási intézményeinek.

Az Ingenta kereskedelmi cég, amelyet azért alapítottak, hogy a tudományos információs lánc valamennyi résztvevőjét összehozzák, beleértve a kiadókon és tudományos társaságokon mint tartalomtulajdonosokon, valamint a kutatókon és a tudományos társaságok tagjain mint végfelhasználókon túl a folyamatba bekapcsolódó közvetítőket is, mint amilyenek a könyvtárak, az előfizetési ügynökségek stb.

Az üzleti modell nagyon egyszerű volt: a tudományos kiadók megfizetik, hogy a nyomtatott kiadványaikban megjelenő tartalmakat pdf vagy html formátumban digitalizálják, és online szolgáltathatóvá teszik az Ingenta szerverén *(www.ingenta.com)*. A könyvtárak és a felhasználók számára a szolgáltatás ingyenes volt, ha a nyomtatott kiadványhoz érvényes előfizetéssel rendelkeztek.

1998-ban a BIDS műszaki rendszerét felhasználva az Ingenta havonta tízezer cikket küldött a folyóiratok előfizetőinek. Ez a szám 2000 márciusára havi százezerre nőtt, amikor az Ingenta megkezdte a felvásárlásokat. Először a CARL Corporation UnCover szolgáltatását vette meg, amely az USA-ban a felsőoktatási szektor legtöbbet használt kereső rendszere volt. Ezzel az Ingenta a világ legnagyobb, tudományos cikkek keresését és szolgáltatását lehetővé tévő forrásának lett a tulajdonosa. A felhasználók által letöltött vagy alkalmilag szolgáltatott (pay-per-view) cikkek száma havonta kétszázezer fölé emelkedett. Az UnCover adatbázis további 21 000 faxon vagy Ariellel küldhető kiadványt tartalmazott, ezáltal az Ingenta belépett a dokumentummásolatot küldő cégek sorába is.

Az első vásárlást tíz hónappal követte a legnagyobb versenytárs, a CatchWorld cég megvétele,